Nucleic Acids Research, 1994, Vol. 22, No. 14 2719–2725

The human DNA polymerase β gene structure. Evidence of alternative splicing in gene expression

Yau-Jan Chyan, Susan Ackerman[2], Nancy S.Shepherd[2], O.Wesley McBride[3], Steven G.Widen[1], Samuel H.Wilson[1] and Thomas G.Wood*

Recombinant DNA Laboratory, Sealy Center for Molecular Science, [1]Sealy Center for Molecular Science, University of Texas Medical Branch, Galveston, TX 77555-0851, [2]DuPont Merck Pharmaceutical Company, Wilmington, DE 19880-0328 and [3]Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Received June 9, 1994; Accepted June 21, 1994 EMBL accession nos U10516–U10526 (incl.)

ABSTRACT

DNA polymerase β (β-pol) is a single-copy gene that is considered to be part of the DNA repair machinery in mammalian cells. Using two human genomic libraries we have cloned the complete human β-pol gene and determined the organization of the β-pol coding sequence within the gene. The human β-pol gene spans 33 kb and contains 14 exons that range from 50 to 233 bp. The 13 introns vary from 96 bp to 6.5 kb. Information derived from this study was used in defining the location of a deletion/insertion type restriction fragment length polymorphism (RFLP) 5′ to exon I of the human β-pol gene. This RFLP was utilized in linkage analysis of DNAs from CEPH families and the results confirm the previous assignment of the human β-pol gene to chromosome 8 (p12–p11). Analysis of mRNA from six human cell lines using the polymerase chain reaction showed the expression of two β-pol transcripts. Sequence analysis revealed that the size difference in these transcripts was due to deletion of the 58 bp sequence encoded by exon II, suggesting that the smaller transcript results from an alternative splicing of the exon II sequence during processing of the β-pol precursor RNA.

INTRODUCTION

DNA polymerase β (β-pol) is the simplest naturally occurring DNA polymerase (1). A monomer of 335 amino acids, this 39 kDa protein has only one enzymatic function and is devoid of such activities as 3′ or 5′ exonuclease, endonuclease, dNMP turnover or the reverse of the DNA synthesis reaction, pyrophosphorolysis (2). β-pol has been implicated in DNA repair (3–6) and appears to be responsible for short gap-filling synthesis (7). Using substrates with gaps of up to six nucleotides, Singhal and Wilson showed that β-pol will completely fill the gap and that the synthesis is conducted in a processive, rather than a distributive manner (7). The processive synthesis has an absolute requirement for a 5′ phosphate at the 5′ end of the gap.

Both the rat and human β-pol proteins have been overexpressed in *E.coli* and these proteins are fully active in DNA synthesis (8–11). Purified recombinant β-pol has been used in controlled proteolytic digestion experiments to create defined polypeptide domains for structural and functional analyses (10, 12, 13). An 8 kDa domain representing the N-terminal 75 residues of β-pol has single-stranded nucleic acid binding activity, but this 8 kDa polypeptide does not bind to double-stranded nucleic acids (ds-NA) and is devoid of nucleotidyltransferase activity. The C-terminal 31 kDa domain has catalytic activity and binds to ds-NA.

The cDNAs for rat and human β-pol have been cloned and sequenced (14, 15). The human β-pol cDNA was used to clone a genomic DNA fragment containing the first two exons of β-pol and the 5′ adjacent flanking region (16). The transcription initiation site and the flanking sequence required for functional promoter activity were mapped using this genomic clone (16). A similar study examining the murine β-pol promoter has also been reported (17). The human β-pol gene has been assigned to chromosome 8 (18), however, the organization of the entire β-pol gene has not been previously reported. We have isolated a set of recombinant DNA clones that contain the entire β-pol gene. The structural organization of the exons encoding β-pol, their sizes and the exon–intron boundaries have been determined.

MATERIALS AND METHODS

Cell culture

All cell lines were grown in a humidified atmosphere at 37°C with 5% CO_2. HeLa cells were propagated as previously described (16). Primary human fibroblast cells (HEMS) were maintained in McCoy's 5a medium containing 4 mM L-glutamine and 25 mM HEPES buffer. Medium was supplemented with 15% heat-inactivated (30 minutes at 56°C) fetal bovine serum, 2 mM L-glutamine, 50 μg/ml penicillin G, 50 μg/ml streptomycin and

*To whom correspondence should be addressed

100 μg/ml neomycin. The normal human colon cell line, CCD18Co (ATCC No. CRL 1459), the colon carcinoma cell line, HCT116 (ATCC No. CCL 247), and the two colon adenocarcinoma cell lines, LS180 (ATCC No. CL 187) and COLO 320 DM (ATCC No. CCL 220) were maintained as recommended by American Type Culture Collection.

Library construction and screening

A human foreskin genomic library (DMPC-HFF # 1) was constructed using a vector system derived from P1 bacteriophage (19–21) and screened by PCR using primers (5′ GAGCTGGGTTGCTCCTGCTC 3′, 5′ GAGCATGTCGGTGATTCCCC 3′) designed from the sequence reported for exon I of human β-pol (16).

A second library was constructed using DNA isolated (22) from a primary human fibroblast cell line (HEMS) derived from embryonic muscle and skin. DNA was partially digested with Sau3AI and size fractionated on sucrose gradients. DNA fragments (14–20 kb) were cloned using EMBL3 and the primary library was amplified using *E.coli* SURE cells (Stratagene). The library was plated using P2392 (Stratagene) and screened using a probe representing the entire human β-pol cDNA coding sequence (11). Probes were labeled using a random primer kit (BRL) and α-^{32}P-dCTP (New England Biolabs) as described (23).

Southern hybridization analysis

Southern blots were prepared as described (24) and hybridizations were performed using 5×SSC, 50 mM sodium phosphate pH6.5, 5 mM EDTA and 0.5% ($^{w}/_{v}$) milk protein at 65°C for 18–20 hours. The PCR primers used in generating a set of hybridization probes representing contiguous 200 bp DNA fragments of the coding region of the human β-pol cDNA are:

001–200:	5′ ATGAGCAAACGGAAGGCGCCGCAGG 3′
	5′ GTTCCTACTCCAGGCAATTTCTTAGCTT 3′
201–400:	5′ AAAAATTGCTGAAAAGATTGATGAGTTTTTAGC 3′
	5′ GGTTCAATTTATCTTCATTTTTTCTGAGATC 3′
402–599:	5′ TCATCAGCGAATTGGGCTGAAATATTTTG 3′
	5′ AAGCTGGGATGGGTCAGGAGAACATC 3′
602–800:	5′ CTTCAGAATCAACCAAACAGCCAAAACTG 3′
	5′ CAGTAATACTGATCTTTGGGTATCAACC 3′
801–1003:	5′ TGGTGTTCTCTATTTCACTGGGAGTGATA 3′
	5′ CGCTCCGGTCCTTGGGTTCCCGG 3′

DNA sequence analysis

Double stranded DNA sequencing reactions were performed using Sequenase version 2 (USB) and cycle sequencing was performed using Taq DNA polymerase (Perkin Elmer). Cycle sequencing reactions were analyzed on an Applied Biosystems model 373A DNA sequence analyzer.

Linkage analysis

DNAs from 40, three-generation CEPH families (25) were analyzed using a probe representing the 5′ 560 bp of human β-pol cDNA. RFLP typing was performed by Southern blot analysis (24) of SstI digested DNA as previously described (18). The entire panel of parental DNAs was initially examined and all family members from informative matings (i.e, one or both parents were heterozygotes) were then genotyped. DNA from some of the families was also analyzed using either BamHI or HindIII digests.

All relevant published markers in the CEPH database v.5 were used for two-point linkage analysis using LINKAGE v.5 for PC compatibles (26). The probe–enzyme combinations for these loci are: D8S87, D8S84, and D8S85 are Mfd39A, Mfd8, and Mfd18 using PCR (27); PLAT is ptFA-4352 with EcoRI; D8S22 is CRI-V1225 with EcoRI; D8S23 is CRI-V822 with EcoRI (28); CAII is H25-3.8 with TaqI; ψJ is pJD0.6 with BglII; D8S2 is 182B with TaqI.

Reverse transcription–polymerase chain reaction (RT–PCR)

Total RNA was isolated (29) and poly A+ RNA selected using oligo-dT cellulose (Pharmacia). First strand cDNA synthesis was performed using 2 μg poly A+ RNA, 50 ng oligo $(dT)_{12-18}$ primer and 500 units of Moloney murine leukemia virus reverse transcriptase (BRL) in a final volume of 50 μl using conditions recommended by the manufacturer. The cDNA was extracted with phenol–choloroform, precipitated with isopropanol and dissolved in 100 μl of water. PCR amplification was performed using 100 μl reaction mixture containing 10 mM Tris–HCl pH8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM each dNTP, 1 μl cDNA, 100 pmoles of each primer and 2–3 units of Taq DNA polymerase (Perkin Elmer). The reaction mixture was subjected to 30 cycles of the following program: denaturation at 95°C for 30 seconds, annealing at 70°C for 30 seconds and primer extension for one minute at 72°C. The final primer extension was extended five minutes to allow for full extension of incomplete products. Amplification products were analyzed by electrophoresis using 5% non-denaturing polyacrylamide gels and visualized by ethidium bromide (1 μg/ml) staining. HaeIII digested φX174 RF DNA (New England biolabs) was included in each gel assay for size estimation of DNA fragments. RT–PCR amplified DNA fragments were cloned into the pCR II vector (Invitrogen) and used in DNA sequence analysis.

RESULTS

Isolation of the human β-pol gene

The cDNA for human β-pol has been cloned and sequenced (11, 15) and a fragment of genomic DNA spanning the first two exons of the human β-pol gene and approximately 11 kb of 5′ flanking region was characterized (16). To isolate the entire human β-pol gene, we utilized two separate libraries. A human foreskin genomic library (DMPC-HFF # 1) was constructed using the P1 bacteriophage cloning system developed by Sternberg and co-workers (19–21). This library was initially screened by PCR using primers (see Methods) that represent sequences derived from exon I of the β-pol gene (16). Three independent clones, DMPC-HFF # 1-0232G03(380), -0570B01(381) and -1150E07(382), were identified from this screening. A second genomic library was constructed using bacteriophage λ EMBL3 and DNA from a primary human fibroblast cell line (HEMS) derived from embryonic muscle and skin. This library was screened (5×10^6 phage) using a probe that represented the entire human β-pol cDNA coding sequence (11). Three independent clones (λ3-1, λ5-1 and λ8-1) were isolated following three rounds of plaque purification.

To facilitate analysis of the human β-pol gene structure, we designed a series of five PCR derived DNA probes that represent contiguous 200 bp segments of the human β-pol cDNA coding region (11). The λ and P1 genomic clones were characterized by hybridization analysis using each of these 200 bp β-pol probes. The λ5-1 clone hybridized exclusively to the probe representing the first 200 bp of the β-pol coding region, while λ8-1 DNA hybridized to probes spanning the 201–800 bp region of β-pol

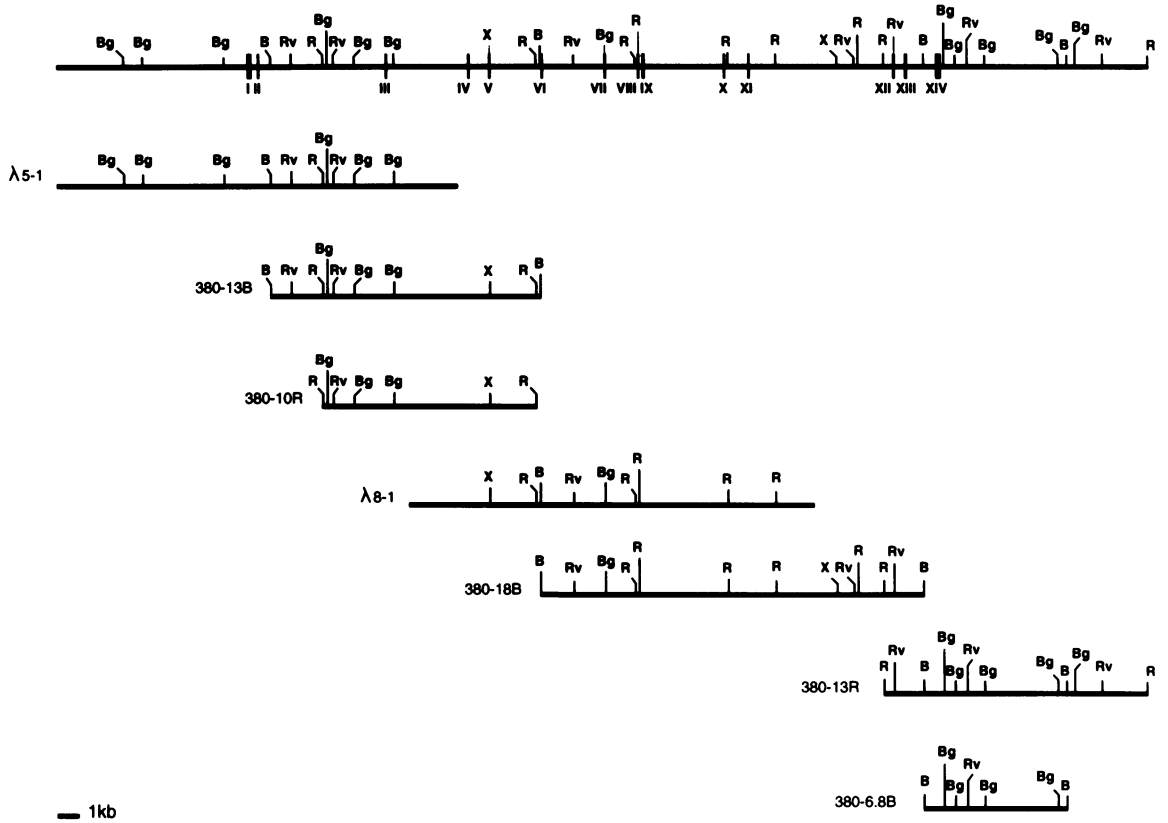

Figure 1. Physical map of the human β-pol gene is shown. Exons are represented as black boxes. A complete restriction map of the region is presented for restriction enzymes: BamHI, B; BglII, Bg; EcoRI, R; EcoRv, Rv and XhoI, X. Plasmid or λ clones used in determining gene structure are shown below the genomic map.

Table I. The sizes of exons and introns in human β-pol gene

Exon	Length (bp)	Amino Acid[1] position	Intron	Type[2]	Length (kb)
I	192	1–20	1	1	0.326
II	58	21–40	2	2	6.0
III	67	41–62	3	0	4.2
IV	75	63–87	4	0	0.8
V	59	88–107	5	2	2.5
VI	50	108–123	6	1	3.0
VII	52	124–141	7	2	1.5
VIII	55	142–159	8	0	0.096
IX	73	160–183	9	1	3.8
X	71	184–207	10	0	1.1
XI	87	208–236	11	0	6.5
XII	65	237–258	12	2	0.513
XIII	140	259–304	13	1	1.5
XIV	233	305–335			

[1]Based upon human β-pol cDNA sequence (11).
[2]Intervening sequence type is defined where type 0 indicates placement between codons, type 1 interrupts a codon between the first and second nt, and type 2 occurs between the second and third nt. of codon (31).

(data not shown). The λ3-1 clone hybridized to probes representing the 400–600 bp region of β-pol. Since this sequence is represented in λ8-1, the λ3-1 clone was not further analyzed in this study. Two of the P1 genomic clones, DMPC-HFF # 1-0232G03(380) and -0570B01(381), hybridized with each of the 200 bp probes, while the third P1 clone did not anneal to the probe representing the last 200 bp segment of the β-pol coding region. Clone DMPC-HFF # 1-0232G03(380) (Clone 380) contained a 62 kb genomic DNA insert and was used in combination with the λ5-1 and λ8-1 clones to determine the structure of the human β-pol gene. Two sub-libraries were generated using the 62 kb insert derived from the P1 clone 380. Insert DNA was digested with either BamHI or EcoRI and then ligated into pBluescriptII KS$^-$ vector (Stratagene). Each sub-library was screened using probe representing the entire coding sequence of the human β-pol cDNA. Five clones, p380-18B,

Table II. Exon–intron boundaries of the human β-pol gene

5'									3'
CTC	ACA	G:	gttagcaccgggccgggccc	intron	1	agccttctgttgcctttcag:	AA	CTC	GCA
	20							22	
GCT	TAC	AG:	gtgggacagtgcagcattct	intron	2	cttttcttctttccttatag:	A	AAA	GCA
	39							41	
AAG	AAA	TTG:	gtaagtttagttagcatgtt	intron	3	atttctaattttccatgtag:	CCT	GGA	GTA
	62						63		
CTG	GAA	AAG:	gtaaaattttaacttgttta	intron	4	gacttttttttttcttaaag:	ATT	CGG	CAG
	87						88		
GGC	ATT	GG:	gtaagaactatttttttaagc	intron	5	taatgcttttgtttttgtag:	T	CCA	TCT
	106							108	
CTA	GAA	G:	gtgagtatgactgtaggtca	intron	6	tttatttatcttctatacag:	AT	CTC	AGA
	123							125	
GGG	CTG	AA:	gtaagatggcagattttctt	intron	7	tacttattctgtctttatag:	A	TAT	TTT
	140							142	
CAA	ATG	CAA:	gtaagatgtgtcaaattata	intron	8	cttaatttttcttctattag:	GAT	ATT	GTA
	159						160		
AGA	AGA	G:	gtaacatacttcctaatctt	intron	9	tgtgtcttctgtcatcacag:	GT	GCA	GAG
	183							185	
ACC	AAA	CAG:	gtgcctcagagtttataatc	intron	10	ccctaattatgattctacag:	CCA	AAA	CTG
	207						208		
AAG	TTC	ATG:	gtaagtacttgttagagtta	intron	11	gtttaaaatgttcattttag:	GGT	GTT	TGC
	236						237		
GAT	ATC	AG:	gtattgttcagactttgttg	intron	12	ctactgtccatttttttttag:	G	TTG	ATA
	257							259	
GTC	ACT	G:	gtgagtgtccatgtgtgtat	intron	13	ttttctctctgtacttgcag:	GA	GTT	GCA
	304							306	

Exon sequence is presented in upper case letters and intron sequence in lower case. The numbers identify codons within the coding sequence of the β-pol cDNA (11).

p380-13B, p380-13R, p380-10R and p380-6.8B, were isolated. All of the clones were characterized by restriction analysis (BamHI, BglII, EcoRI, EcoRV and XhoI), and the DNA fragments were analyzed by Southern blot hybridization using probes representing one or more of the 200 bp segments derived from the human β-pol cDNA coding region. DNA fragments shown to contain exon sequences were used in a detailed restriction mapping, and DNA sequence surrounding the exon–intron boundaries was determined. The results of this analysis are summarized in Figure 1.

Exon–intron organization of the human β-pol gene

The human β-pol gene can be constructed using five overlapping recombinant clones (Fig 1). Two additional clones (p380-10R and p380-6.8B) were useful in delineation of the restriction map. The human β-pol gene is organized into 14 exons that span a total of 33 kb (Fig. 1). The exons range in size from 50 to 233 bp (Table I). Exon I and II were previously described (16). Exon I (192 bp) contains 61 bp of coding sequence and 131 bp of 5′ untranslated sequence. Exon XIV, the largest exon (233 bp), contains 92 bp of coding sequence prior to a TGA termination codon and 138 bp of 3′ untranslated sequence. The largest exon with respect to coding sequence for β-pol, is exon XIII (140 bp). Introns vary from 96 bp (intron 8) to approximately 6.5 kb (intron 11). All exon–intron boundaries have a high degree of homology with the 5′ and 3′ splice junction consensus sequences (30) (Table II), and each of the three types of exon–intron junctions (31) are represented within the human β-pol gene (Table I). Table II provides sequence information for each of the exon–intron boundaries. The complete cDNA sequence has been published (11, 15) and additional information regarding the sequence of human β-pol introns can be accessed from GenBank (U10516–U10526 inclusive).

Table III. RFLP at β-pol locus*

	SstI	HindIII	BamHI
A1	6.6	8.2	17.5
A2	11.3	14	2.7
Invariant	24	8.9	14

*Sizes (kb) of allelic (A1, A2) and invariant restriction fragments detected with 560 bp β-pol 5′ cDNA probe are shown. Allele frequencies were determined in 39 unrelated NIH blood donors (A1:A2=0.86:0.14) and in the 80 CEPH parents (A1:A2=0.90:0.10).

Linkage analysis of human β-pol gene

A deletion/insertion type restriction fragment length polymorphism (RFLP) is present in the 5′ flanking region of the human β-pol gene (18). This RFLP occurs in approximately 20% of the population and is detected in several different restriction digests (Table III). Restriction mapping of the 5′ region of the β-pol gene using SstI or HindIII suggests that the RFLP may result from an insertion or duplication of approximately 4–6 kb (Table III). The presence of a 2.7 kb BamHI allele combined with the absence of any detectible difference in human DNA digested with BglII (data not shown) suggests that the deletion/insertion occurs within approximately 0.5 kb 5′ of the first BglII site (Fig.1) preceding exon I.

The genotypes of a few informative CEPH families were previously examined using this RFLP (32). All 80 CEPH parents and each member of the 14 informative families have been genotyped using this RFLP and used for linkage analysis with other loci on chromosome 8 (Table IV). The results indicate a close linkage between β-pol, the other loci on proximal 8p (D8S87, PLAT and D8S22) and the centromeric region of chromosome 8. There were two recombinants (141803 and

Table IV. Two-point Lod scores for β-pol with other loci on chromosome 8

Locus[a]	Θ[b]	Z[c]	Confidence[d] interval	Physical location
D8S87	0.055	13.7	0.020−0.140	8p12
PLAT	0.00	19.0	0.00−0.036	8p12−q11.2
D8S22	0.00	11.4	0.00−0.040	
D8S23	0.22	2.3		
CAII	0.21	1.7		8q13−q22
ψJ	0.11	1.1		
D8S84	0.23	3.5	0.11−0.35	8q13−q21.2
D8S2	0.22	1.2		
D8S85	0.63	0.3		8q21−q22

[a]Loci are listed in the most probable order based on multipoint linkage analysis (32). See Method for probe−enzyme combinations. All genotypic data except β-pol and ψJ was generated in other laboratories: J.C.Weber (D8S87, D8S85), E.Robson (PLAT), S.Wood (CAII, D8S2), T.Jenkins (D8S84), and H.Donis-Keller (D8S22, D8S23).
[b]Most likely recombination fraction between β-pol and each locus.
[c]Most likely Lod scores assuming no sex difference in recombination frequencies.
[d]Confidence intervals for recombination fractions over a 10-fold range of likelihood.

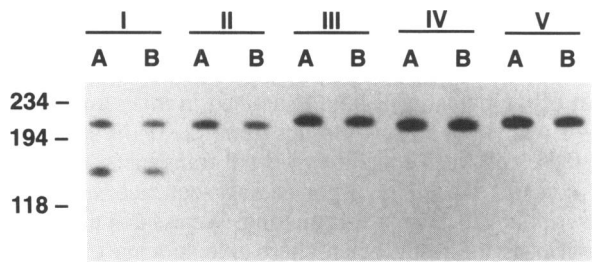

Figure 2. Analysis of amplified DNA from RT−PCR. Poly A+ RNA from either HeLa (A) or HEMS (B) cells was isolated and used in RT−PCR assays (Methods). The primers used in this experiment were designed based upon the sequence of the coding region of the human β-pol cDNA (11) and represent a set of contiguous target DNAs (200 bp) that collectively span the entire cDNA. Primer pair I, 1−200; II, 201−400; III, 402−599; IV, 602−800 and V, 801−1003.

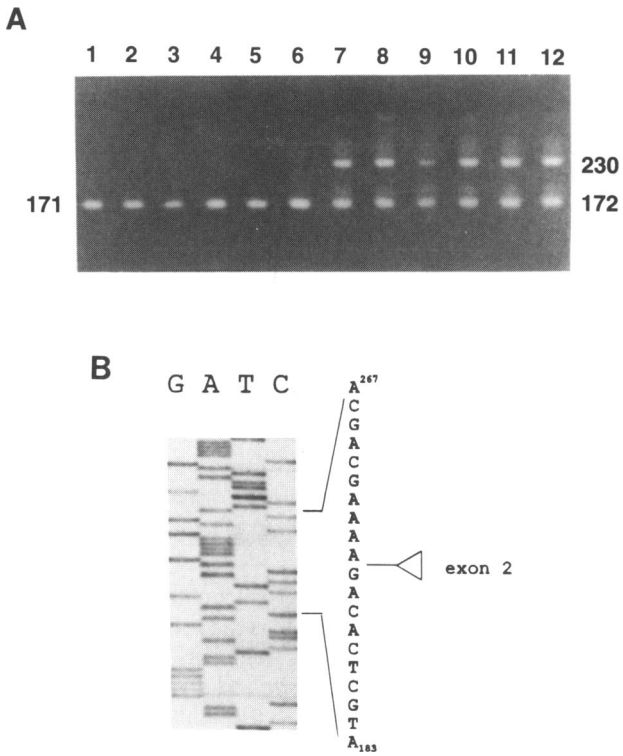

Figure 3. RT−PCR amplification and DNA sequence analysis of human β-pol mRNA. **A.** Poly A+ RNA was isolated from each cell line and used in an RT−PCR assay. Primers used in this assay lanes 1−6, S1 and AS1 lanes 7−12, S1 and AS2.
S1 5′ AGTCCTGGTACCTCCTTCAAGCTG 3′
AS1 5′ GCTCACGTTCTTCTCAAAGTTTGCGAGTT 3′
AS2 5′ GGGTATTTTGCTATAACAGATGCTGCTTTT 3′
Lanes 1 and 7, HeLa mRNA; lanes 2 and 8, HEMS mRNA; lanes 3 and 9, CCD 18 Co mRNA; lanes 4 and 10, LS180 mRNA; lanes 5 and 11, COLO 320 DM mRNA and lanes 6 and 12, HCT 116 mRNA. **B.** DNA sequence analysis. The 172 bp DNA fragment (HeLa mRNA) was cloned into pCR II (Invitrogen) and used in DNA sequence analysis. DNA sequence for the junction between exon I and III is shown. Identical results were obtained for each analysis using plasmid DNA from five independent clones.

141804) between D8S87 and the three adjacent loci (β-pol, PLAT and D8S22), indicating that D8S87 is probably distal (telomeric on 8p) to both β-pol and PLAT. Since there are no recombinants between β-pol and PLAT or D8S22, we cannot order β-pol with respect to these loci. PLAT and D8S22 are very closely linked with a single recombinant between them. This genetic localization of β-pol is consistent with the previously reported localization of the gene on chromosome 8 (p12−p11) (33, 34).

Analysis of β-pol transcripts

The size of β-pol mRNA expressed in human cells has been previously reported as ~1.4 kb (14, 15). We have examined several human cell lines using Northern blot hybridization analysis and found identical results in every cell line tested regardless of the region of the β-pol cDNA that was used as the hybridization probe (date not shown). However, small differences in transcript size (≤ 100 bp) would not be easily detected in Northern blot assays, and since the size of eleven of the β-pol exons is less than 100 bp, we decided to examine the β-pol mRNA using a reverse transcriptase−polymerase chain reaction assay (RT−PCR). In this assay, we utilized the same primers as were employed in generating the set of 200 bp probes representing contiguous segments of the entire β-pol cDNA coding region. Poly A+ RNA isolated from both HEMS and HeLa cells was used in the RT−PCR assay (Figure 2). The anticipated PCR product was obtained with both HeLa and HEMS poly A+ RNA using each of the five primer pairs (Figure 2, I−V). DNA fragments of a size other than the predicted target DNA were observed only with primer pair I. With this primer pair, two DNA fragments (200 bp and 142 bp) were amplified in the RT−PCR reaction, suggesting the presence of two different β-pol mRNAs. The first 200 bp of the coding sequence of β-pol are distributed over exon I, II, III and the first 14 bp of exon IV (Table I). To determine the sequence deleted in the 142 bp DNA fragment, we analyzed both HeLa and HEMS poly A+ RNA using a sense primer representing sequences derived from the 5′ untranslated region of exon I (S1) and an antisense primer from the immediate 5′ end of either exon II (AS1) or exon III (AS2) in an RT−PCR

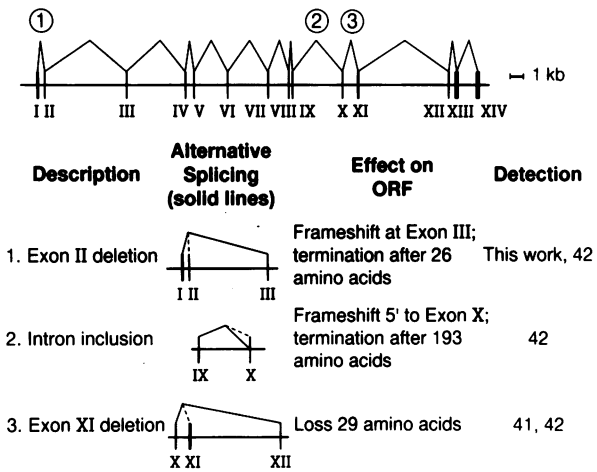

Figure 4. Summary of putative alternative splicing of β-pol mRNA.

assay (Figure 3A). The 171 bp product obtained using the exon II-specific antisense primer (AS1) is the predicted size of the β-pol target sequence. Two DNA fragments, a 230 bp fragment representing the predicted target size and a 172 bp fragment, were amplified using the exon III-specific antisense primer. DNA sequence analysis of the 172 bp DNA fragment revealed that the 58 bp sequence encoded by exon II was absent (Figure 3B). The sequence of the 230 bp fragment was identical to the predicted target region of the β-pol cDNA (data not shown).

A similar experiment was performed using poly A+ RNA isolated from a normal human colon cell line (CC8-18Co), a human colon carcinoma cell line (HCT116), and two human adenocarcinoma cell lines (COLO320DM and LS180). RT−PCR results using either the 200 bp primer pairs (data not shown) or the exon II- and exon III-specific antisense primers (Figure 3A) were identical to those obtained using HeLa or HEMS poly A+ RNA. Each cell line contains a population of β-pol specific RNA that does not contain sequences present in exon II. This was the only difference detected in the β-pol transcripts expressed in these cells.

DISCUSSION

DNA polymerase β is a constitutively expressed DNA polymerizing enzyme (35, 36) that is conserved throughout vertebrates (15, 37, 38) and has enzymatic properties consistent with a role in DNA repair (3−7). In this report, we show that the gene encoding the 39 kDa β-pol protein is organized into 14 exons that extend over 33 kb of genomic DNA. The β-pol exons range in size from 50 to 233 bp while the introns vary in size from 96 bp to approximately 6.5 kb. The exon−intron boundaries utilized in the splicing events that generate the 1.4 kb β-pol mRNA are highly homologous to the 5′ and 3′ splice junction consensus sequences (30). A 114 bp sequence 5′ to the β-pol transcription start site has been shown to have promoter activity (16, 39, 40). This promoter does not contain CAAT or TATA box sequences but does have three decanucleotide elements that have homology with the Sp1 transcription factor binding site and a decanucleotide sequence starting at position −40 that has perfect dyad symmetry and matches the binding site for the ATF/CREB transcriptional activator (16, 39, 40). The transcription initiation site is 131 bp 5′ to the ATG initiation codon for β-pol. This 5′ untranslated sequence and the first 61 bp of the coding region of β-pol comprise exon I. The next eleven exons (II−XII) range in size from 50 to 87 bp, while the last two exons, XIII and XIV, consist of 140 bp and 233 bp, respectively. Exon XIV contains only 92 bp of β-pol coding sequence. The remaining sequence following a TGA termination codon represents 138 bp of 3′ untranslated sequence.

Approximately 20 percent of the human population carry a deletion/insertion type restriction fragment length polymorphism (RFLP) that is detected with probes containing sequences derived from the 5′ end of β-pol (18). Results from restriction analysis suggest that this RFLP most likely results from a 4∼6 kb insertion or duplication that maps approximately 1.6 kb 5′ to exon I of the β-pol gene. Linkage analysis using this RFLP and the 80 CEPH families confirmed earlier reports for the localization of β-pol gene on human chromosome 8p12−p11 (33, 34). It should be noted that three recombinants were found between PLAT and both β-pol and D8S87 in a single family (142404, 142406, 142410), whereas no recombinants were found between β-pol and PLAT in any of the other families. Unfortunately, D8S22 and other loci were not informative in this family. We concluded that the most likely explanation for these results was genotyping errors in PLAT and these three genotypes were excluded from the analysis shown in Table IV.

All of the human cell lines examined in this study contained an ∼1.4 kb β-pol mRNA. Results from RT−PCR analysis show that these cell lines contain two β-pol transcripts. One mRNA that contains the entire β-pol coding sequence and a second transcript in which exon II is missing. We assume that this latter transcript is the result of an alternative splicing event (Figure 4). The omission of exon II alters the open reading frame resulting in a termination codon after the first 26 amino acids. The first 20 amino acids, which are encoded in exon I, are identical to residues in the β-pol protein, while the last six amino acids (glu-lys-gln-his-leu-leu) are the result of a frame-shift with the polypeptide terminating at a TAG codon. A functional role for either the alternative splicing pathway or the potential protein product(s) generated by this splicing event requires further investigation.

Wang *et al.*, examined β-pol mRNA expressed in human colorectal tumor specimens (41). While they did not observe exon II deletions in their assays, they did detect several different deletions in β-pol mRNA from tumor tissue that were not found in the normal mucosal samples. The predominate deletion observed in their study was an 87 bp sequence that encodes amino acid residues 208−236 (41). This corresponds exactly to the region of β-pol encoded by exon XI (Figure 4). The exon−intron junctions for both intron 10 and intron 11 interrupt the coding sequence between codons (Table II). Therefore, a deletion of exon XI would not alter the β-pol reading frame for sequences downstream from, exon XI.

Sodakane *et al.* recently reported on β-pol mutations observed in patients with Werner's syndrome (42). A 107 bp insertion was detected in β-pol transcripts using an RT−PCR analysis of RNA isolated from two unrelated patients with Werner's syndrome. Our data suggest that the reported 107 bp insertion results from aberrant splicing of the primary β-pol RNA, not from an insertion (Figure 4). The 107 bp addition occurs at the junction between exons IX and X (Table I). We find an identical sequence (107 bp) within intron 9, 92 bp from the 5′ end of exon X. Intron sequences 5′ and 3′ to this 107 bp sequence are compatible with

splice acceptor (TGTTGTCATCACAG·A) and splice donor (AA·GTAAGG) consensus signals (30). The cause of the aberrant splicing event, its frequency within the human population, and its significance relative to Werner's syndrome, is not clear at this time. A mutation corresponding to the 87 bp deletion reported in human colorectal tumors (41) was also found in one of the patients with Werner's syndrome (42). These investigators did observe the 58 bp deletion representing sequences encoded in exon II in all of the human samples analyzed.

The structural information provided in this report regarding exon organization, exon−intron boundaries, and DNA sequence will be of significant value to future analysis of the role of β-pol in colorectal cancer, Werner's syndrome, and in the process of DNA repair.

ACKNOWLEDGEMENTS

We wish to thank Ms Maribel Acosta for her excellent technical assistance, Mr Chinnappa Kodira for help in DNA sequence analysis and Ms Manuella Barnard for her efforts in preparing the manuscript.

REFERENCES

1. Wilson, S.H. (1990) In Strauss, P.R. and Wilson, S.H. (eds.), The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assemblies. Teleford Press, Caldwell, N.J., Vol.I, pp. 199−233.
2. Wilson, S.H., Singhal, R.K., and Kumar, A. (1992) In Bohr, V.A., Wassermann, K. and Draemer, K.H. (eds.), DNA Repair Mechanisms. Munksgaard, Copenhagen, Vol. 35, pp. 343−360.
3. Dresler, S.L. and Lieberman, M.W. (1983) *J. Biol. Chem.* 258, 9990−9994.
4. Smith, C.A. and Okumoto, D.S. (1984) *Biochemistry* 23, 1383−1390.
5. DiGiuseppe, J.A. and Dresler, S.L. (1989) *Biochemistry* 28, 9515−9520.
6. Hammond, R.A., McClung, J.K. and Miller, M.R. (1990) *Biochemistry* 29, 286−291.
7. Singhal, R.K. and Wilson, S.H. (1993) *J. Biol. Chem.* 268, 15906−15911.
8. Abbotts, J., SenGupta, D.N., Zmudka, B., Widen, S.G., Notario, V. and Wilson, S.H. (1988) *Biochemistry* 27, 901−909.
9. Date, T., Yamaguchi, M., Hirose, F., Nishemoto, Y., Tanehara, K. and Matsukage, A. (1988) *Bicochemistry* 27, 2983−2990.
10. Kumar, A., Abbotts, J., Karawaya, E. and Wilson, S.H. (1990) *Biochemistry* 29, 7156−7159.
11. Patterson, T.A., Little, W.A., Wilson, S.H., Widen, S.G., Cheng, X., Kumar, A. and Beard, W.A. (1993) GenBank accession no. L11607.
12. Kumar, A., Widen, S.G., Williams, K.R., Kedar, P., Karpel, R.L. and Wilson, S.H. (1990) *J. Biol. Chem.* 265, 2124−2131.
13. Casas-Finet, J.R., Kumar, A., Morris, G., Wilson, S.H. and Karpel, R.L. (1991) *J. Biol. Chem.* 266, 19618−19625.
14. Zmudzka, B.Z., SenGupta, D.N., Matsukage, A., Cobranchi, R., Kumar, P. and Wilson, S.H. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5106−5110.
15. SenGupta, D.N., Zmudzka, B.Z., Kumar, P., Cobianchi, F., Skowronski, J. and Wilson, S.H. (1986) *Biochem. Biophys. Res. Commun.* 136, 341−347.
16. Widen, S.G., Kedar, P. and Wilson, S.H. (1988) *J. Biol. Chem.* 263, 16992−16998.
17. Yamaguchi, M., Hirose, F., Hayashi, Y., Nishimoto, Y. and Matsukage, A. (1987) *Mol. Cell. Biol.* 7, 2012−2018.
18. McBride, O.W., Zmudzka, B.Z. and Wilson, S.H. (1987) *Proc. Natl. Acad. Sci. USA* 84, 503−507.
19. Sternberg, N. (1990) *Proc. Natl. Acad. Sci. USA* 87, 103−107.
20. Pierce, J.C. and Sternberg, N. (1992) *Methods Enzymol.* 216, 549−574.
21. Shepherd, N.S., Pfrogner, B.D., Coulby, J.N., Ackerman, S.L., Vaidyanathan, G., Sauer, R.H., Balkenhol, T.C. and Sternberg, N. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2629−2633.
22. Wood, T.G., McGeady, M.L., Blair, D.G. and Vande Woude, G.F. (1983) *J. Virol.* 46, 726−736.
23. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6−13.
24. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503−517.
25. Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J.M. and White, R. (1990) *Genomics* 6, 575−577.
26. Lathrop, G.M., Lalouel, J.M., Julier, C. and Ott, J. (1984) *Proc. Natl. Acad. Sci. USA* 81, 3443−3446.
27. Weber, J.L., Kwitek, A.E., May, P.E., Patterson, D. and Drabkin, H. (1990) *Nucl. Acids Res.* 18, 4038.
28. Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S., Botstern, D., Akots, G., Rediker, K.S., Gravius, T., Brown, V.A., Rising, M.B., Parker, C., Powers, J.A., Watt, D.E., Kauffman, E.R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T.R., Ng, S., Schumm, J.W., Braman, J.C., Knowlton, R.G., Barker, D.F., Crooks, S.M., Lincoln, S.E., Daly, M.J. and Abrahamson, J. (1987) *Cell* 51, 319−337.
29. Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* 162, 156−159.
30. Mount, S.M. (1982) *Nucleic Acids Res.* 10, 459−472.
31. Sharp, P.A. (1981) *Cell* 23, 643−646.
32. NIH/CEPH Collaborative Mapping Group. (1992) *Science* 258, 67−86.
33. Cannizzaro, L.A., Bollum, F.J., Huebner, K., Croce, C.M., Chung, L.C., Xuq, X., Hecht, B.K., Hecht, F. and Chang, L.M.S. (1987) *Cytogenet. Cell Genet.* 46, 589.
34. McBride, O.W., Kozak, C.A. and Wilson, S.H. (1990) *Cytogenet. Cell Genet.* 53, 108−111.
35. Chang, L.M.S., Brown, M. and Bollum, R.J. (1983) *J. Mol. Biol.* 74, 1−8.
36. Mitchell, J., Karawya, E., Kinsella, T. and Wilson, S.H. (1985) *Mutat. Res.* 146, 295−300.
37. Chang, L.M.S., Plevani, P. and Bollum, F.J. (1982) *Proc. Natl. Acad. Sci. USA* 79, 758−761.
38. Tanabe, K., Yamaguchi, T., Saneyoski, M., Yamaguchi, M., Matsukage, A. and Takahashi, T. (1984) *J. Biochem.* (Tokyo) 96, 365−370.
39. Widen, S.G. and Wilson, S.H. (1991) *Biochemistry* 30, 6296−6305.
40. Narayan, S., Widen, S.G., Beard, W.A. and Wilson, S.H. (1994) *J. Biol. Chem.* 269, 12755−12763.
41. Wang, L., Patel, U., Ghosh, L. and Banerjee, S. (1992) *Cancer Res.* 52, 4824−4827.
42. Sodakane, Y., Maeda, K., Kuroda, Y. and Hori, K. (1994) *Biochem. Biophys. Res. Commun.* 200, 219−225.